DOCUMENT RESUME

ED 295 999 TM 011 912

AUTHOR Anderson, Paul S.
TITLE Comparison of Student Attitudes about Seven Formats of Educational Testing, with Emphasis on the MDT Multi-Digit Testing Technique.
PUB DATE 87
NOTE 38p.; Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Chicago, Illinois, October 15-17, 1987). For a related document, see TM 011 913.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Students; Comparative Analysis; *Computer Assisted Testing; Essay Tests; Higher Education; Multiple Choice Tests; Objective Tests; *Scoring; *Student Attitudes; Test Anxiety; Test Construction; *Test Format; Test Items
IDENTIFIERS *MDT Multiple Digit Testing; Short Answer Tests

ABSTRACT

Seven formats of educational testing were compared for student test preferences and how well each evaluated learning. The formats were: (1) true/false; (2) multiple choice; (3) matching; (4) MDT Multiple Digit Testing, in which a machine scores fill-in-the-blanks; (5) fill-in-the-blanks; (6) short answers; and (7) essay. A total of 1,440 survey questionnaires (the Survey of Student Opinions about Methods of Educational Testing) completed at a Midwestern university by students exposed to the MDT method obtained information about student opinions. The MDT was not as well received as were more familiar tests. Students thought themselves less able as test takers. When the sample was controlled for appropriateness to the class, the MDT compared with fill-in-the-blanks for evaluative power and preference. The machine-scored test is practical for teacher use. Eleven figures are presented, and an appendix includes the survey. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Comparison of Student Attitudes about Seven Formats of Educational Testing, With Emphasis on the MDT Multi-Digit Testing Technique

ED295999

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paul S. Anderson
Geography-Geology
Illinois State University
Normal, Illinois 61761

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
PAUL S. ANDERSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Seven formats of educational testing are compared on student test preferences and perceptions of how well each test method evaluates learning: 1) True/False, 2) Multiple-Choice, 3) Matching, 4) MDT Multi-Digit Testing, 5) Fill-in-the-Blank, 6) Short Answers, 7) Essay. MDT Multi-Digit Testing is a machine-scored equivalent to fill-in-the-blank tests. It utilizes numerically labeled, alphabetized long lists of up to 1000 discrete responses. A survey of 1440 college students reveals that students perceive a hierarchy in the formats of educational testing. The above list ranks them in increasing complexity of responses, increasing student perception of ability to evaluate learning, and decreasing student preference. The newly devised MDT Multi-Digit Test is not as favorably received by students as are more familiar methods. Students consider themselves to be less able as test takers with the MDT method. Students indicate no familiarity with it from high school and comparatively little from university courses. Thirty-five percent stated that the MDT method was not appropriate as used in their course. When the sample was controlled for "appropriateness," the MDT method was as well liked and with equal evaluative power as the fill-in-the-blank method. The incorporation of the more rigorous MDT method of evaluation into the upper realms of machine-scored testing should benefit education in terms of learning and savings of time and costs.

[A paper presented at the annual conference of the Mid-Western Educational Research Association (MWERA) in Chicago, IL., on 15-17 October 1987.]

TM011912

A. Introduction:

This study compares seven formats of educational testing in terms of two key issues that focus on student attitudes. The first is student preference for certain types of testing, that is, which test formats students like and dislike. The second is student perception of how well each test method evaluates student learning. The seven test methods [and their abbreviations as used in this study] are given below. They are in the rank order of the general complexity of their responses in terms of the number of alternative responses from which the students are to formulate their answers.

1. [T/F] True/False (dichotomous responses)
2. [MC] Multiple-Choice (usually five alternate responses for each question; the responses can be phrases; students are expected to read all alternatives).
3. [MAT] Matching (short lists of responses, usually fewer than 20 foils shared by several question stems; students usually read all of the foils).
4. [MDT] MDT Multi-Digit Testing (long list of up to 1000 discrete alphabetized responses; list can be long to discourage searching to recognize a response).
5. [FIB] Fill-in-the-Blank (infinite mental bank of discrete, free responses)
6. [SA] Short Answers (one or two sentence responses in free format)
7. [ESS] Essay (paragraph or longer responses in free format)

The first three formats can be machine-scored while the last three are manually scored. All six of those six methods are widely used and familiar to students. The middle method, MDT multi-digit testing, is less well known because it became available only in the mid-1980s.

B. Explanation of the MDT Technique:

The MDT multi-digit testing method is essentially a machine-scored "fill-in-the-blank" test. Technically, the MDT technique is all of the following: machine-scored, clued free-response, discrete answer, multiple-digit, and long-list answer bank educational testing, with distinctive computer assisted processing and feedback.

The stems of the questions are prepared in a normal manner. As shown in Figure 1, an example would be: "Name the second president of the United States." Students who know the answer look at a provided alphabetized long-list to obtain the associated label number. The label number is marked on a machine-readable answer sheet. The students who do not know the answer are unable to select the correct label because the list (or "answer bank") with up to 1,000 discrete alternatives is intentionally too long to allow searching for unknown answers. Those who know the answer (John Adams in this example) will easily find the code number in the "A" section of the MDT list. Much more thorough descriptions and discussions of the technique are in The MDT Innovation (Anderson, 1987a).

The multi-digit testing technique has been used since 1983 with over eight thousand student enrollments at Illinois State University and has recently been introduced at several other schools. The MDT method is applicable to all fields of study at all educational levels from upper elementary through graduate school, including training programs and competency testing. Physicians are expected to know certain facts about anatomy and medicine, while seventh grade students are expected to know facts appropriate to their grade level. Instructors retain complete control of the content covered and the question difficulty, as with regular fill-in-the-blank testing.

The MDT testing technique is not a research instrument in this study. Rather, it provides the "treatment" about which the students express their attitudes. The method is examined in this research in its hypothesized role as an intermediate be-

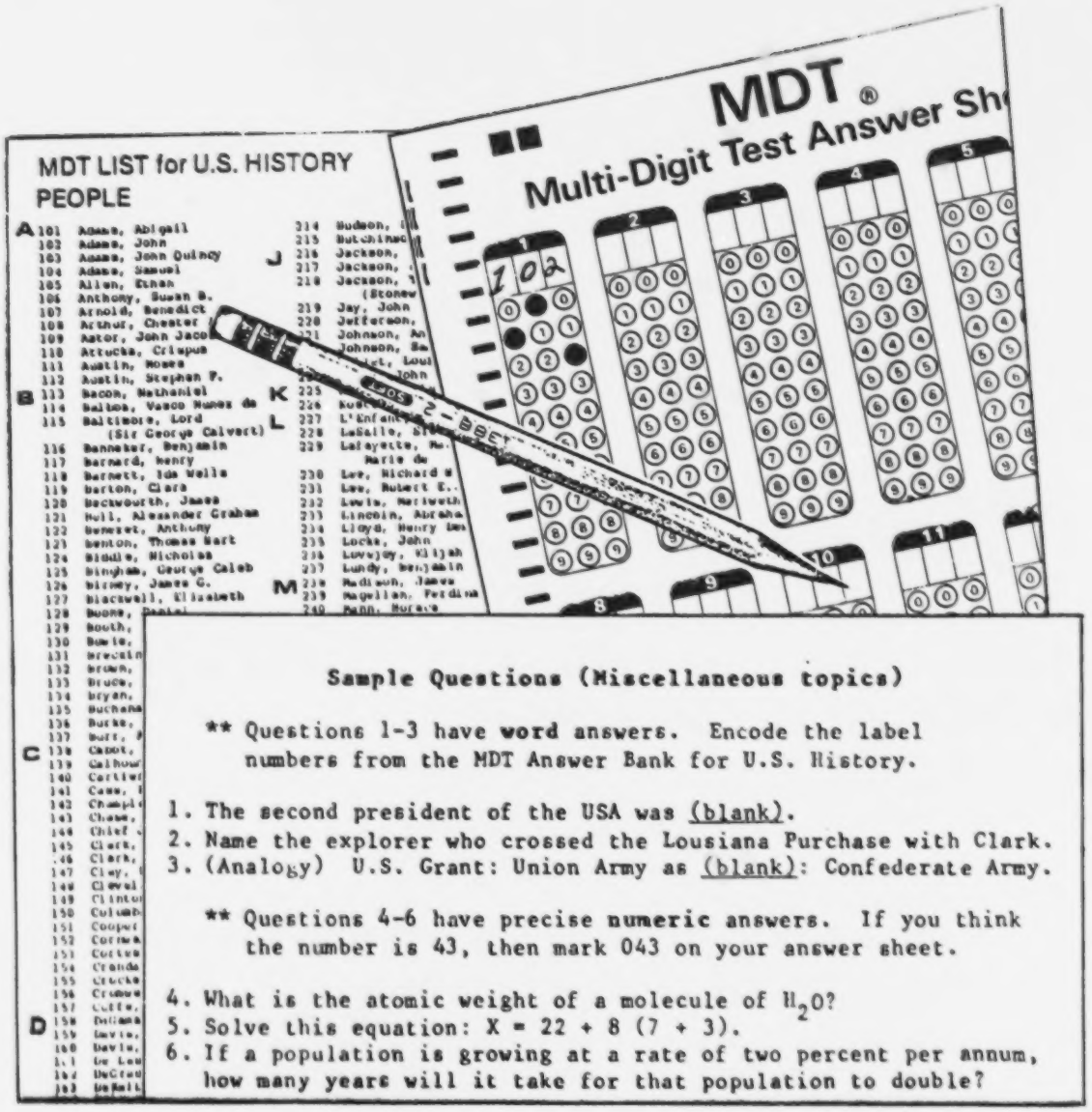
MDT LIST for U.S. HISTORY
PEOPLE
MDT ®
Multi-Digit Test Answer Sh
Sample Questions (Miscellaneous topics)
** Questions 1-3 have word answers. Encode the label numbers from the MDT Answer Bank for U.S. History.
1. The second president of the USA was (blank).
2. Name the explorer who crossed the Lousiana Purchase with Clark.
3. (Analogy) U.S. Grant: Union Army as (blank): Confederate Army.
** Questions 4-6 have precise numeric answers. If you think the number is 43, then mark 043 on your answer sheet.
4. What is the atomic weight of a molecule of H_2O?
5. Solve this equation: X = 22 + 8 (7 + 3).
6. If a population is growing at a rate of two percent per annum, how many years will it take for that population to double?

Figure 1: Examples of MDT Multi-Digit Testing materials, including questions, "answer bank" list and MDT answer sheet.

BEST COPY AVAILABLE

tween multiple choice and fill-in-the-blank test styles. The effect of "appropriate" usage is examined.

C. Data Source, Methods and Initial Analyses:

At a reasonably typical Midwestern university with over 20,000 students, twenty sections of students in diverse courses were exposed to the MDT method as part of the educational testing during the Fall 1986 semester. An end-of-semester "Survey of Student Opinions about Methods of Educational Testing" was collected from those students (see Appendix A). A total of 1440 completed questionnaires constitute response from 80% of all students tested by the MDT method in that semester. However, the instructors in classes were not a random sample of all university courses. Therefore the results cannot be applied to student bodies with different attributes.

The questionnaire included 58 variables for student characteristics and opinions plus one variable to identify each of the 20 courses. Included in this questionnaire were five sets of seven questions dealing with the seven formats of educational testing being evaluated. The first set (A or HS-EXP) asked how much experience did the student respondents have with these testing methods in their high school education. The second set (B or UNIV-EXP) was similar, but with reference to their university level education. The third set (C or TT-ABLE) asked the students to rate their ability as test takers with each of those seven testing methods. The fourth set (D or EVAL) asked the students to rate the test methods according to how well each method could evaluate student learning. Finally, the fifth set (E or GENATT) asked "In general, what is your attitude about each testing method?" Each of the seven questions in the five sets was rated with a semantic differential on a scale of 1 thru 5.

For each of those five major sets of responses on the questionnaire, each of which included reference to the seven testing methods, the overall average (mean)

response was calculated when taking all of the seven testing methods into account at the same time. For example, in Set A (Ques. 21-Ques. 27) where the students comment on their experience with the testing methods in their high school education, each student's response values from 1 through 5 for each of the seven testing methods were added together. The possible range of values was from a minimum of 7 through a maximum of 35 if the student answered all of the questions. That value was divided by the number of responses given, thereby obtaining the average value per student. These were summed and divided by the number of students to determine the overall averages. It is acknowledged that the student responses are on an ordinal scale. Therefore the average values are at best an approximation for the responses of the students. These averages for the 1440 respondents are in the columns marked TOT (for total sample) in Figure 2. These values are discussed later in this paper.

It is also possible to analyze how each student respondent views the variety of test formats. Since the responses in themselves refer to a wide range of tests from true/false through essay, and since it is likely that students do have preferences and different levels of experience, it is expected that the average score should tend to be fairly uniform and close to the middle value of three. This was found to be the case, with four of the five means ranging from 3.2 to 3.6 with standard deviations ranging from .48 to .58. The one exception was the mean for university experience (2.9 with standard deviation .65). This slight inconsistency is attributed to the fact that 50 percent of the sample were first-semester college freshmen who therefore had not yet had a significant number of courses to have experienced a wide variety of testing methods at the university level.

Since these newly calculated means tend toward the central values, it was not unexpected that those five composite variables yielded almost no statistically significant nor noteworthy correlations with the personal characteristics of the students in Questions 1 through 18 on the questionnaire. The only instances where the correlation coefficients exceeded .20 were with reference to Set C (test taking

	Attitude (E)			Rate How Well Evaluates (D)			Ability as Test Taker (C)			Experience in University (B)			Experience in High School (A)		
	TOT	APP	INAP	TOT	APP	INAP	TOT	APP	INAP	TOT	APP	INAP	TOT	APP	INAP
1 T/F	3.21	3.11	(3.37)	2.67	2.65		3.44	3.39		3.16	3.17		3.99	3.99	
2 MC	3.87	3.87		3.40	3.39		3.75	3.75		4.33	4.38		4.35	4.36	
3 MAT	3.55	3.56		3.30	3.31		3.73	3.75		2.66	2.67		3.82	3.83	
4 MDT	2.49	2.89	(1.74)	3.12	3.46	(2.49)	2.86	3.18	(2.27)	2.37	2.48	(2.15)	1.35	1.32	
5 FIB	2.87	2.88		3.79	3.86	(3.68)	3.16	3.20	(3.09)	2.45	2.44		3.63	3.64	
6 SA	3.17	3.14		4.04	4.06		3.49	3.49		2.66	2.66		3.69	3.71	
7 Essay	2.99	2.93		4.12	4.14		3.50	3.47		2.74	2.73		3.50	3.53	

TOT = total
APP = appropriate
INAP = inappropriate

Figure 2: Table of mean values of student responses to five questions about each of seven test formats for the total sample (N=1440) and the "appropriate" subsample (N=921) [Note: Unless shown in parentheses, the means for the "inappropriate" subsample are virtually the same as those for the other two means.]

ability) which correlated with the following student attributes: a) "overall grade point average" (Ques. 6) with a value of r = 0.287; b) Ques. 9 where the students give a self rating of their "natural intelligence (ability)" (r = 0.363); and c) correlations of r = 0.271 and 0.225 with "expected grade" and "deserved grade" (Questions 10 and 11, respectively). In other words, the academically stronger students considered their abilities to take tests in a full range of formats to be greater than did less strong students.

Apart from the above mentioned correlations, there is evidence that the five composite sets of variables are basically independent of the individual characteristics of the students.

The survey results were tabulated and processed. Where appropriate, Pearson analyses were used to identify correlations between variables. In the cases of discrete variables, ANOVA was utilized to identify statistically significant differences between the mean values.

In an earlier paper (Anderson, 1987b), the research focused on student attitudes toward the MDT method. The first conclusion derived from the data was that of the seven test formats, all except the MDT method had a near-normal distribution of student attitudes (See Figure 3, which shows the results of Questions 49-55. Those questions constitute Set E, for which the mean values are given in Figure 2.) In a bimodal distribution, thirty percent of the 1440 respondents gave the least favorable ("strongly dislike") rating as their attitude in Question 52 about the MDT format of testing..

Five variables (Questions 52, 56, 57, 58 and 59) were combined to formulate a composite dependent variable of attitude (ATT) toward the MDT method (see Figure 4). The ATT variable correlated highly with each of the five source variables (the range of Pearson's r was from 0.7791 to 0.9081 (see Figure 5).

That earlier research revealed that the two independent variables with the highest correlations with favorable student attitudes ATT were Ques. 16 ("Are the

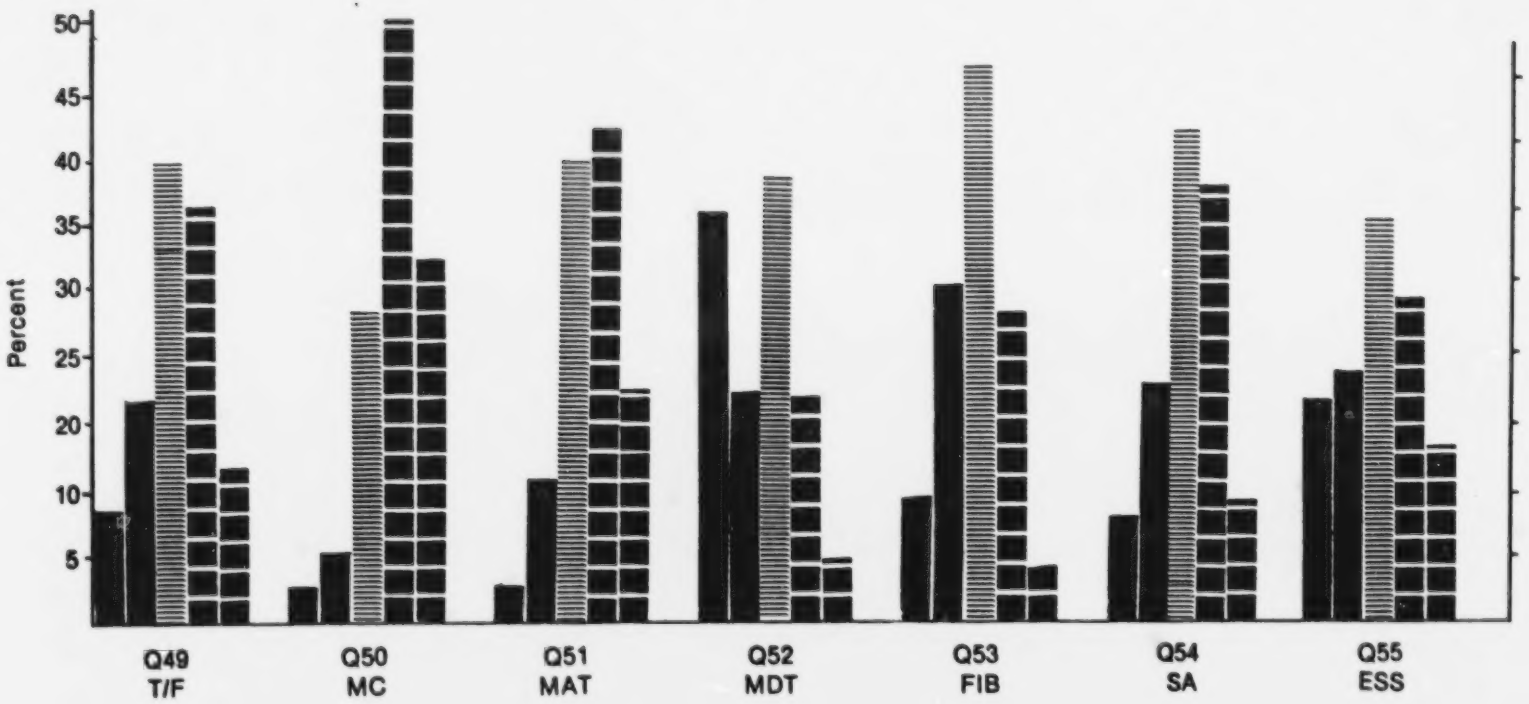

Figure 3: Histograms of student attitudes about disliking or liking each of seven test formats. (Whole sample, N=1440).

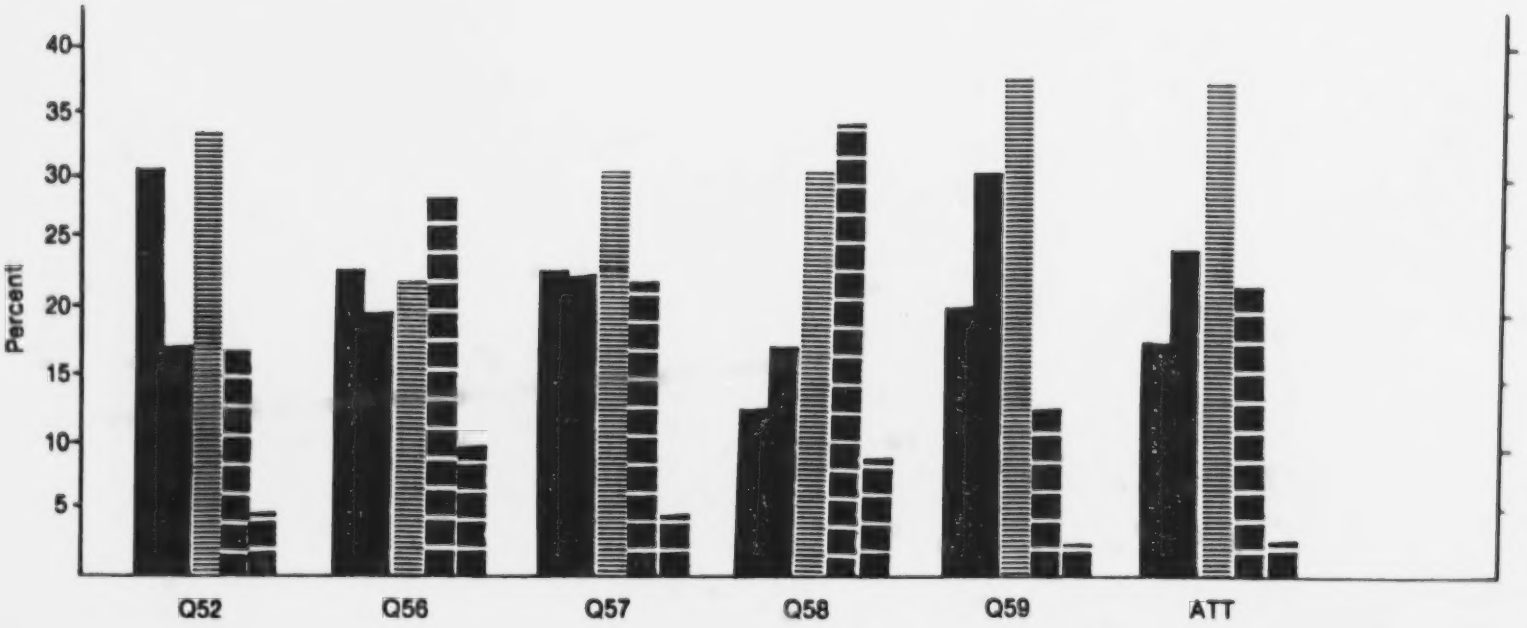

Figure 4: Histograms of five expressions of student attitudes about the MDT multi-digit testing format, plus the composite ATT attitude variable. (Whole sample, N=1440).

ATT	-----					
52	.8461 (.8178)					
56	.9081 (.8683)	.7165 (.6412)	----			
57	.8617 (.8339)	.6530 (.5763)	.7380 (.6931)	----		
58	.7791 (.6502)	.5214 (.3699)	.6528 (.4798)	.5864 (.4232)	----	
59	.8385 (.7771)	.6789 (.6018)	.7107 (.5837)	.6695 (.5791)	.5358 (.3380)	----
	ATT	52	56	57	58	59

Figure 5: Pearson correlation coefficients between the five variables (Questions 52, 56, 57, 58, 59) that are combined into the dependent variable of attitude (ATT) toward the MDT method. [p<0.0000__ in all cases] Upper values are for the entire sample (n=1400+); Lower values in parentheses are for the subsample that considered the MDT method to be appropriately used in the course (n=900+).

MDT testing procedures as used in this course appropriate for the course material?") and Ques. 7 to "rate the instructor". The correlation coefficients were r = 0.639 and r = 0.349, respectively. Those two variables (Ques. 16 and Ques. 7) were only correlated to each other at the value of r = 0.294, indicating that the two are not simply mirrors of each other and that they can be used jointly for analyses of course-related influences upon the attitudes toward the MDT and other test formats.

D. Analysis of Course-Related Influences

The research reported in this present paper attempts to control for course-related influences upon attitude. A third variable, Question 18 concerning the fairness of the grading in the course as perceived by the student, was also added. The three variables were combined by taking the mean scores for each student for those three variables and forming a derived variable called "bad experience" (BADX).

Upon computation of the BADX derived variable, a dichotomous split was made at the mean value of less than or equal to 2.0 out of 5.0. Six of the twenty course sections in the survey had high percentages of students indicating a "bad experience". Those percentages were from 15.8 up to 26.1 percent. None of the other fourteen classes was above 8.5 percent, with the average being only 3.2 percent. Those six classes were temporarily removed from the sample.

Upon calculation of new values of the student attitudes (ATT) concerning the MDT method, the removal of the six classes with "bad experience" produced only a relatively minor shift toward making the student attitudes about the MDT method approximate a more normal distribution. The interpretation was that the derived variable of "bad experience" was insufficiently precise to be used as a control or filter for the data.

An analysis was made of only Question 16 (appropriateness of the MDT method in the course) which was the single most highly correlated variable. Tallies revealed that five course sections had high percentages of students indicating the very inap-

propriate or inappropriate categories. Those high percentages ranged from 47.3 percent up to 73.5 percent. The other 15 classes had percentages of 34.0 percent or lower, the lowest being 5.0 percent. Interestingly, only three of those five courses were also among the six courses identified in the bad experience (BADX) composite variable discussed above. In other words, two new classes were added in and three other classes were returned to the more normal categories. Essentially, the variables on rating the instructor (Ques. 7) and commenting on the fairness of the course grading (Ques. 19) were clouding the issue concerning the student attitudes toward the MDT method. When calculcations were made of the ATT attitude variable and the fifteen sections that had high percentages of students indicating the appropriateness of the MDT method for that course, it was found that the distribution of student attitudes (ATT) about the MDT method was approaching a normal curve, but that there were still relatively high percentages of students in the lowest categories. The interpretation was that by eliminating these five classes, and likewise when the six classes were separated in the "bad experience" analysis, the net effect was to remove many other students who did not have unfavorable attitudes toward the MDT method, while concurrently leaving in the remaining course sections numerous students who indicated they felt that the MDT method was inappropiate for the subject matter.

Question 16 allowed for four response levels ranging from "very inappropriate" to "very appropriate" use of the MDT method in the course. Ogives were drawn for each of thoses four levels to show the cumulative percentages of students at each of the calculated attitude levels of the ATT variable concerning the MDT method, as shown in Figure 6. The data as graphed indicate that there would be an appropriate division between levels 1 and 2 on the one hand and levels 3 and 4 on the other. The levels 3 and 4 ("appropriate" and "very appropriate") combine to form an almost normal curve of student attitudes toward the MDT format, as illustrated in the central graph in Figure 7.

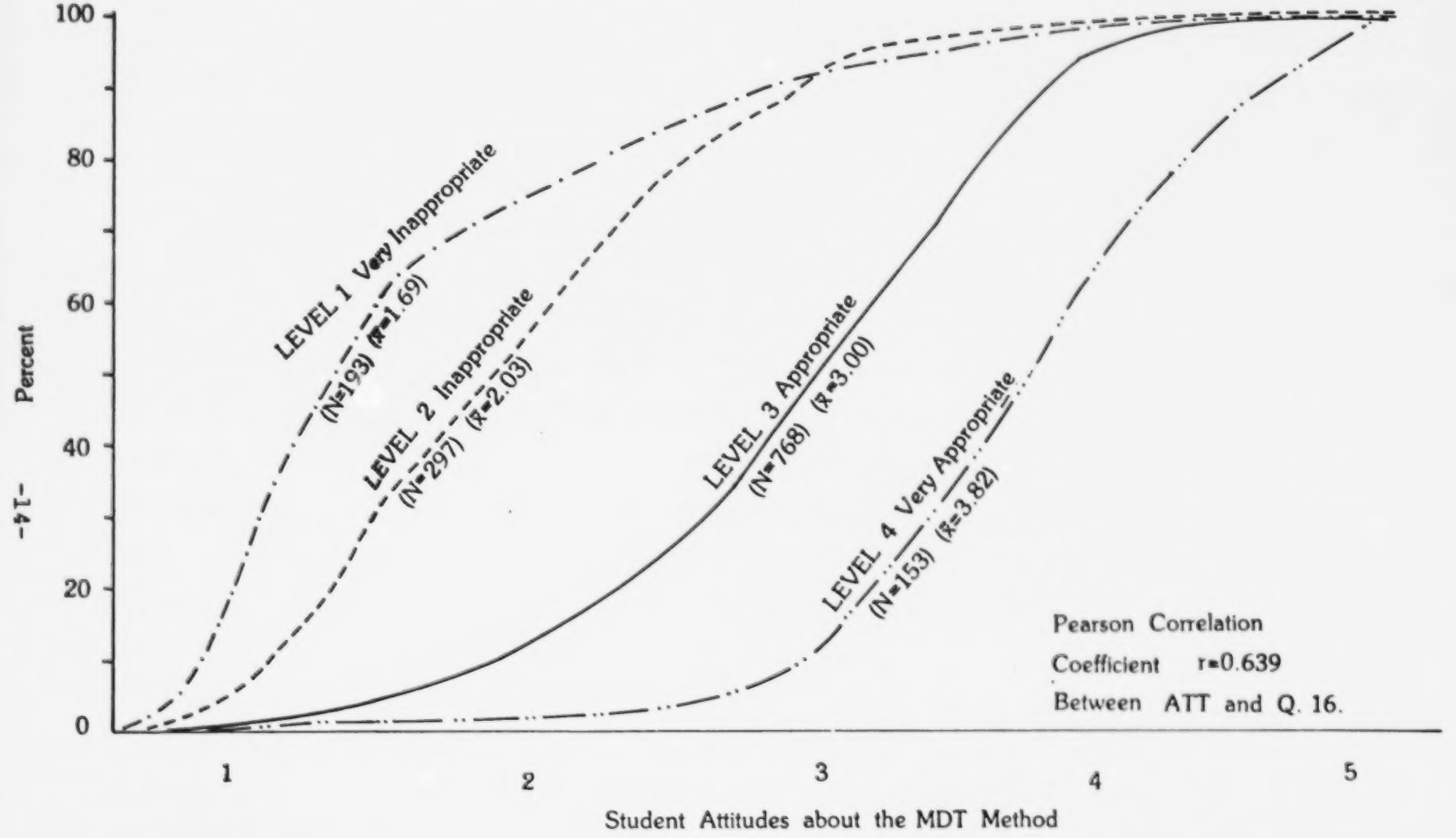

Figure 6: Ogives of student attitudes (ATT) about the MDT method for four levels of appropriateness (Question 16). [Note: Left tails less than 1 are extrapolations.]

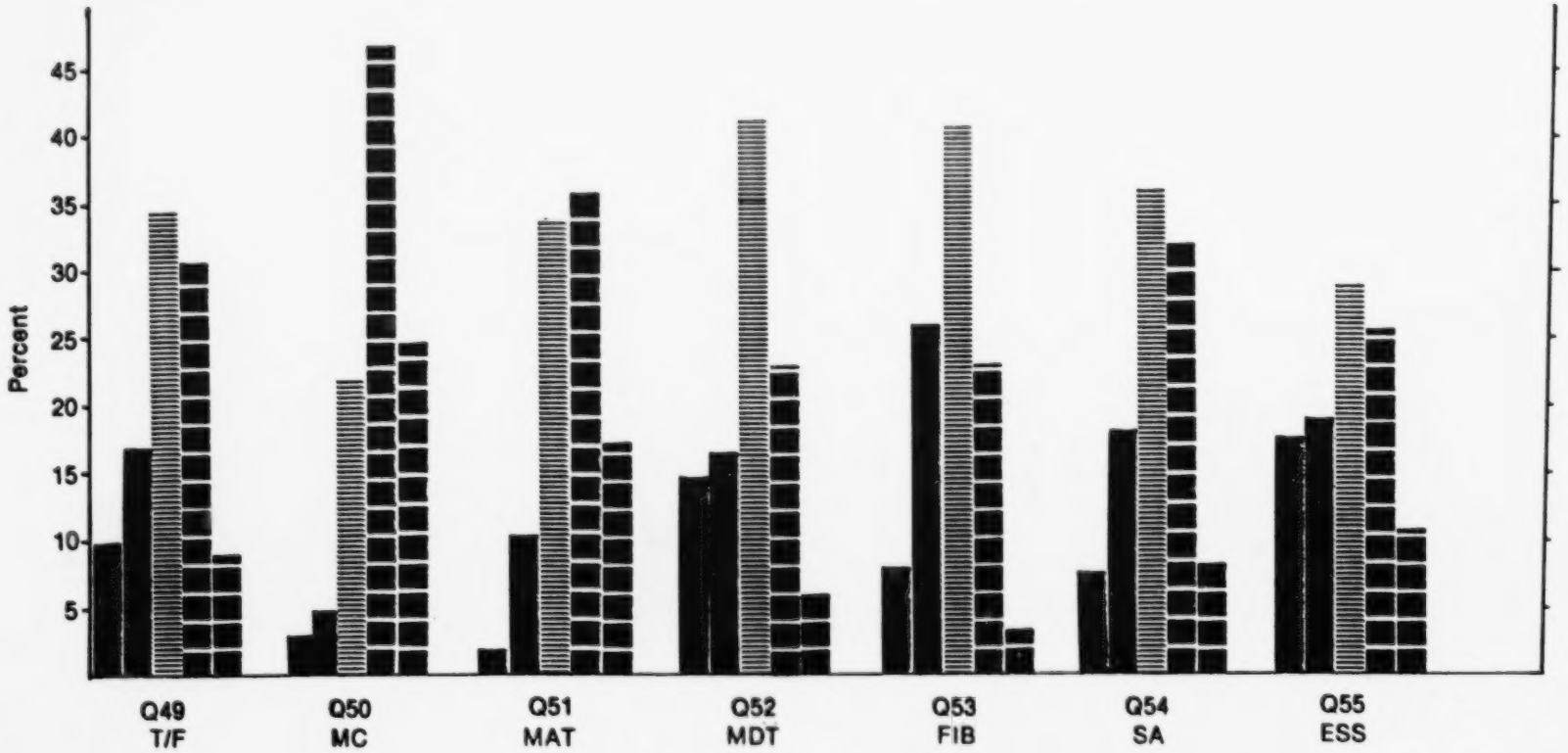

Figure 7: Histograms of student attitudes about disliking or liking each of seven test formats. ("Appropriate" subsample, N=921) (Compare with Figure 3.)

E. Rationale for the "Appropriate" Subsample:

Based on the above data and arguments, it was decided to analyse a subsample that contained only those students who indicated that the MDT method, as used in their class, was appropriate or very appropriate. The rationale for this decision is not on the basis of sampling technique, but on the basis of having a subsample which is representative of what is expected when this MDT test method is used appropriately. "Appropriateness" is a complex issue which has at least three major factors. One factor is how the instructor utilizes the method in the classroom. In this research it was impossible to control each of the instructors in terms of the styles of questions written with the MDT method. Nor was there control over the amount of explanation of the MDT method given by the instructors to their students. In other words, an instructor who is unclear with his or her course objectives and/or is inconsistent with the usage of this or any other testing method for evaluating those course objectives is essentially "evaluating inappropriately" and would receive such a comment from the students on a survey questionnaire.

Second, it is possible that some subject matter included in the tests was not appropriate for the MDT method. Determining what is and is not appropriate in each of the many disciplines is an issue which will require time and care to refine. It is reasonable to expect in the not too distant future that experienced instructors will not use the MDT method in instances where it is indeed inappropriate.

Third, it is also reasonable to expect that students who feel that it is inappropriate might change their minds in the future when they are more familiar with the method. For example, students absent during the explanation of the testing method could subsequently be caught by surprise by the rigor of this new machine-scored testing technique. It would be very natural for some of those students to complain and blame the inappropriateness of the method. This relates to the issue of the "newness" of the MDT method.

To a large extent, Question 16 relating to "appropriateness" is a surrogate measure for the "newness" of the testing method. Newness can be a factor with: a) inappropriate use for certain subject matter, b) insufficient experience and preparation on the part of instructors, or c) a lack of familiarity with the method on the part of the students. In any combination, the issue of newness is highly suggestive of the issue of appropriateness. Therefore, it is reasonable to expect in the future that relatively few students would continue to respond that the MDT method was inappropriate in their course. This might well require several years of experience. But that perceived appropriateness is as likely to occur for the MDT testing format as it has obviously occurred for the multiple choice and other testing formats in America. For the most part all of the testing formats are well understood and properly used by both instructors and their students.

As a test of the reasonableness of the preceding paragraphs, there should not be appreciable differences in the characteristics of the students who indicated that the MDT method was inapproprate in comparison with the characteristics of those who said that it was appropriate. This is indeed the case. Even more important, the separation of the "appropriate" subsample yields no noteworthy changes in the student attitudes toward the other six test formats. See Figure 7 and compare it with Figure 3. The evidence is that there is no difference of meaningful consequence to this research between the students who have been separated out and those who remain in the subsample, those being the students who indicated that the MDT method is either appropriate or highly appropriate for the course in which they were enrolled.

F. Analyses of the Seven Test Methods:

The analyses which follow are based on data derived from the "appropriate" subsample described above. Most explicitly, the subsample includes students who consider the MDT method to have been used appropriately in their course in which they had exposure and then subsequently responded to the questionnaire. It is assumed

that the formulation of the subsample is a reasonable and sufficiently fair step in the analysis process to allow the MDT method to be included into the hierarchy of testing with the other six test methods. It is important to note that Question 16, which was the basis for construction of the subsample, is not a dependent variable used in the formulation of the composite attitude variable called "ATT". Nor does Question 16 eliminate from the analyses the influence of the instructor and the characteristics of the students.

As shown by the numbers in the parentheses in Figure 5, there are still strong correlation coefficients between the five variables used to define the ATT independent variable of student attitude. The coefficients reveal that the influence of the subdivision using Question 16 for appropriateness has resulted in a reduction of the coefficients in all cases. Histograms of the response frequencies for each of the five dependent variables and the composite dependent variable ATT are shown in Figures 4 and 8 for the total sample and the "appropriate" subsample respondents, respectively. The impact of the division according to appropriateness is quite notable. Mean values for the "appropriate" subsample have raised approximately 0.5 units. The subsample is considerably more positive concerning these variables. For purposes of contrast, the negative feelings expressed by those students in the "inappropriate" group are typified by the mean values of approximately 1.7 for those variables.

How much the sampled students like or dislike each of the seven methods of testing are shown in Figures 4 and 7. After control for the issue of appropriateness of use in the classes (Ques. 16), the MDT method is quite similar to that of the fill-in-the-blank style of test questions. Neither of those two methods is particularly well liked, being in the same category as essay questions. It is somewhat surprising to note how favorably the short answer questions are considered, although matching and multiple choice are far more highly liked by the students.

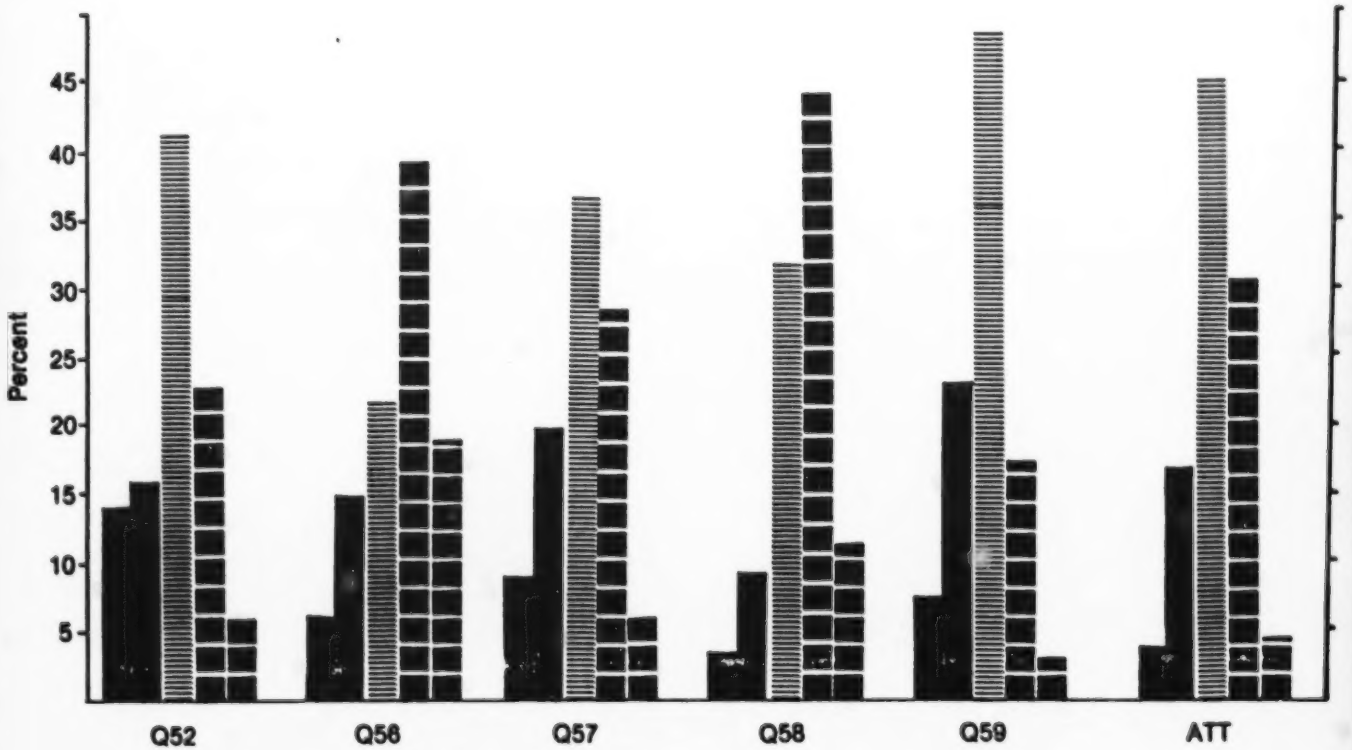

Figure 8: Histograms of five expressions of student attitudes about the MDT multi-digit testing format plus the composite ATT attitude variable. ("Appropriate" subsample (N=921)); (Compare with Figure 4.)

Histograms of how students rate the seven testing methods in terms of their ability to evaluate learning are in Figure 9 for the total sample and Figure 10 for the "appropriate" subsample. Four observations are important. First, in all cases except that of the MDT method, it is most notable how the total sample and the subsample are similar in their responses about how well the different test formats evaluate student learning. Second, in the case of the MDT method, the shift is pronounced in a positive direction for the "appropriate" subsample. Third, in terms of mean values as summarized in column D of Figure 2, the MDT method fits precisely into its hypothesized position between that of the fill-in-the-blank method and that of the machine-scored techniques. Fourth, the means decline steadily from essay at the highest end (see also Figure 11). We note how substantially lower the true/false method is with regard to students perceptions of how well it evaluates learning.

G. Results and Conclusions:

On the basis of mean scores, there are distinct hierarchies in the ratings of the seven test formats in terms of student attitudes about preference and ability to evaluate. Most notable is an inverse correlation between the two data sets (Figure 11). Of the seven traditional testing methods, multiple choice ranks about average (3.4 on a 5.0 scale) on the student perception of its ability to evaluate learning. However, it is the highest in the other four sets of questions. Interestingly, the fill-in-the-blank method is third highest in ability to evaluate learning (3.8), but is rated lowest in the other four sets. Essays are among the least liked but are rated highest in evaluation ability. It is concluded that in a general body of American college students, the testing methods which are perceived to be the best evaluators of learning are the test methods that the students like the least. This does not mean that students object to being well evaluated, but it does indicate a preference for easier methods, that is, methods for which there are fewer responses

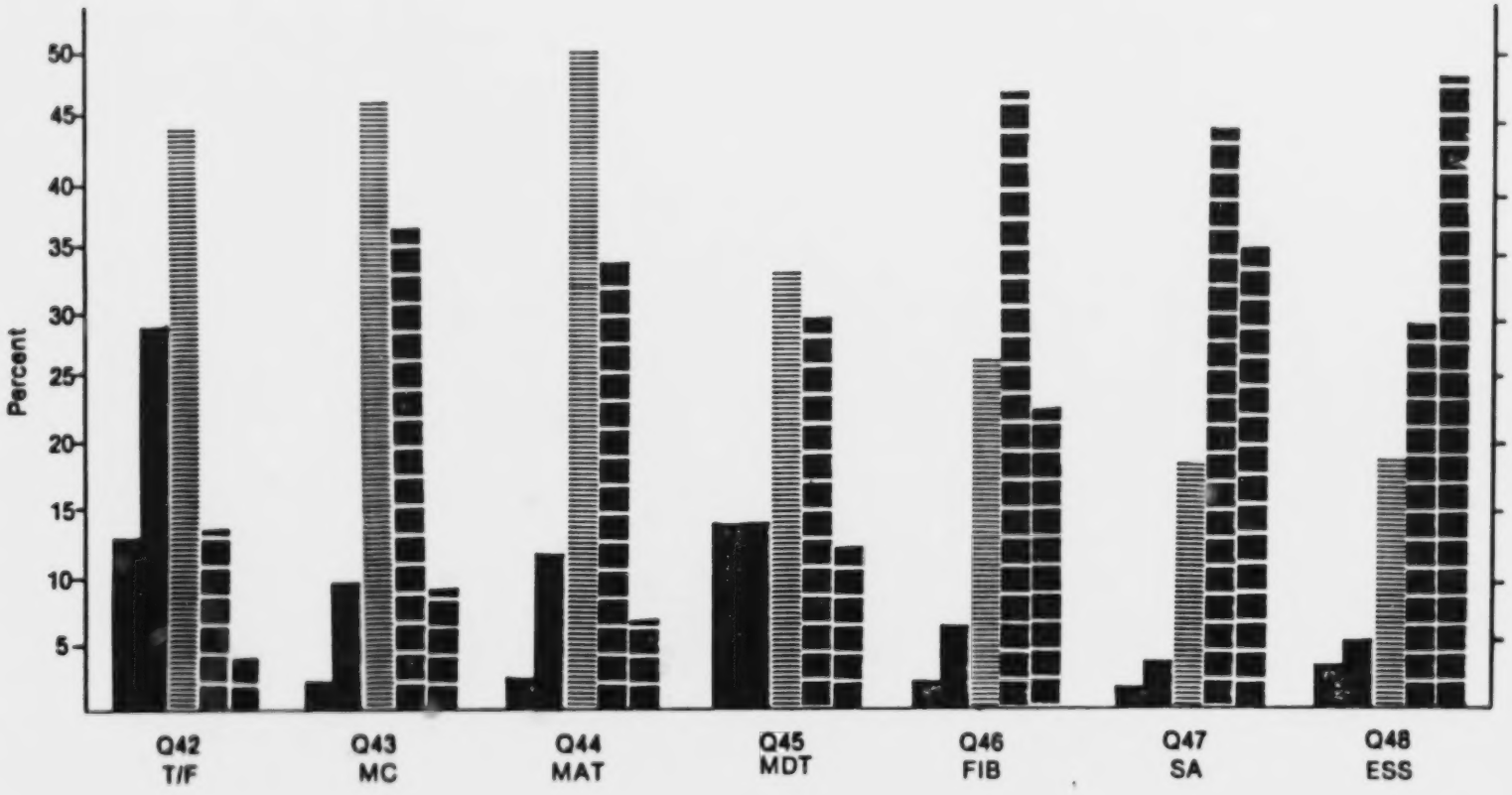

Figure 9: Histograms of student opinions of how well each of seven test formats evaluates student learning. (Total sample, N=1440)

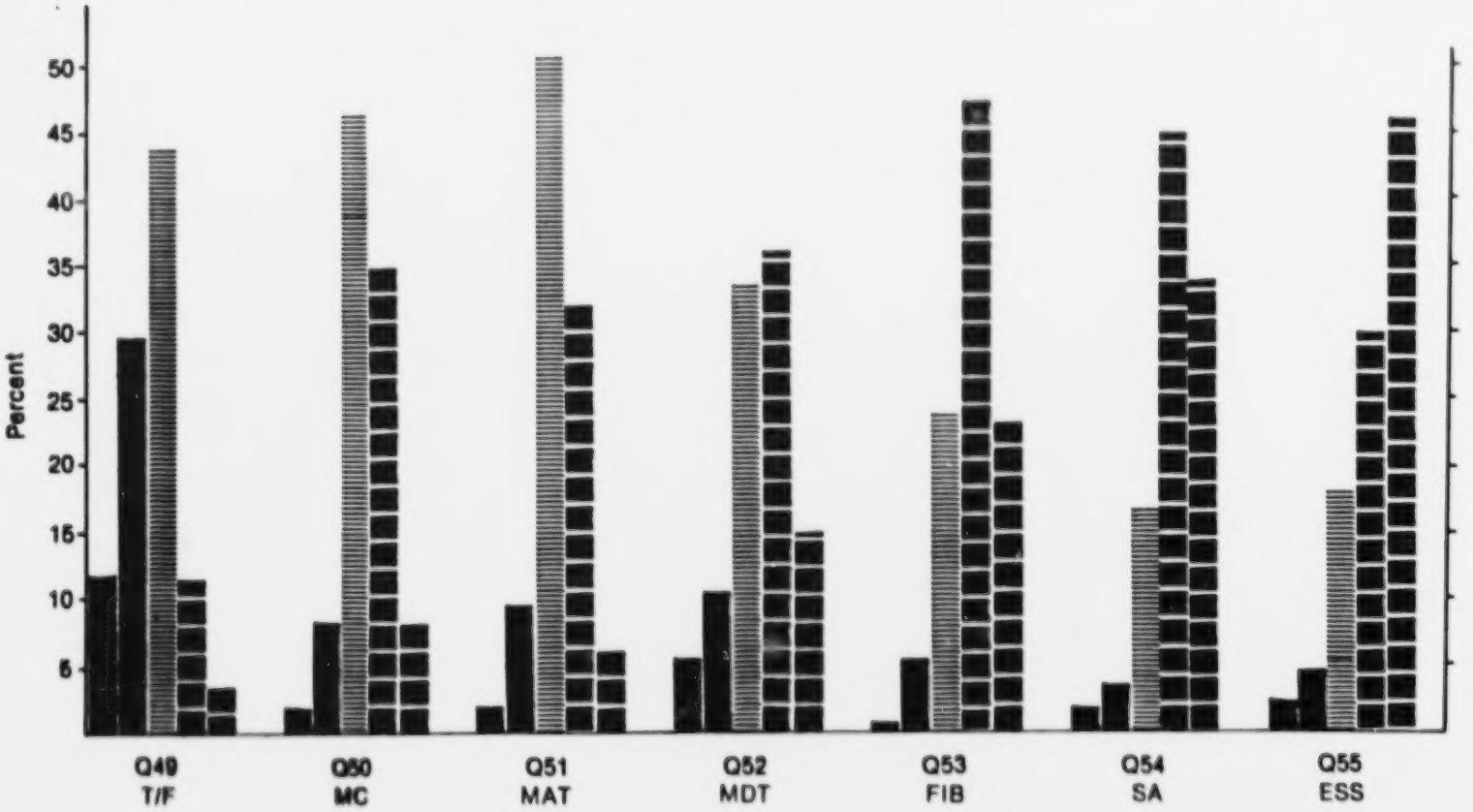

Figure 10: Histograms of student opinions of how well each of seven test formats evaluates student learning. ("Appropriate" subsample, N=921)

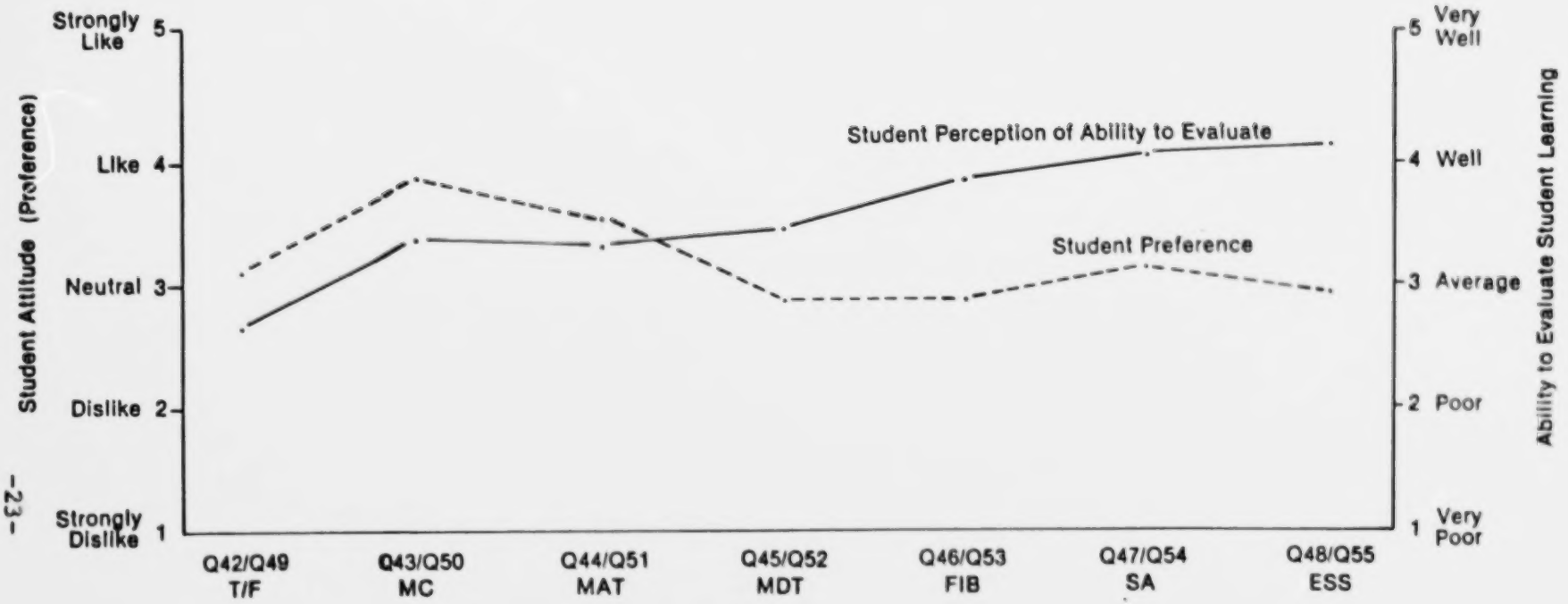

Figure 11: Graph of mean values of student attitudes (like) and perceptions of evaluation ability of seven test formats. ("Appropriate" subsample, N=921)

from which to choose.

The ratings of the MDT method by the "appropriate" subsample fit as expected into the hierarchy. the MDT technique does appear to be a bridge between machine-scored and free-response methods of evaluation. In the perceptions of students, the MDT method is similar to the fill-in-the-blank style of testing. In terms of the students' ability as test takers, they consider themselves to be less able with the MDT method (see table.) In turn, ability is partially a function of prior experience with the MDT testing method. The data sets A and B in the table in Figure 2 reveal that the students have virtually no familiarity with the MDT method from their high school experience, and comparatively little experience from their university courses. Analyses still in progress are attempting to control for this lack of familiarity and to then see how familiarity impacts upon the students' stated attitudes and ratings of the MDT test method.

Analyses of the complete data set reveal that the second highest correlate with the composite ATT dependent variable of "attitude toward MDT testing" was how the students rated their instructor. With $r = 0.349$ and $p = 0.000$_, the impact of the instructor upon student attitudes toward MDT testing is noteworthy. By controlling for instructor-related factors, more uniformity of the subsamples can assist in the study of the student-related variables. It is anticipated that ongoing analyses will indicate that the rating of the MDT method will become more similar to that of the fill-in-the-blank testing method, the technique with which the MDT method is most similiar.

H. Educational Importance of the Study:

The improvement of education in America is in part dependent upon the increase in academic rigor in the educational courses offered. In this country, where any individual of reasonable competence can enroll in some institution of higher education, the percentage of young adults (ages 18-22) enrolled is extremely high. One

outcome from this opening of the doors and opportunities for higher education has been an increasing reliance upon machine-scored testing. Although such methods have their limitations, they are widely accepted because of a substantial body of research that couples nicely with the time and financial benefits of machine scoring. On the other hand, as indicated by the student opinions about the testing methods, those machine-scored methods are rated lower as a means of evaluating students' learning.

The incorporation of more rigorous methods of evaluation into the machine-scored realm has been a dream of many researchers and educators. However, efforts to incorporate the "free response" nature of essays and short answers and even fill-in-the-blank questioning have been fought with frustrations. The MDT method is specifically designed to be a machine-scored alternative for fill-in-the-blank style questions. In its present format and based upon in-class experiences, it appears to successfully fill that niche. Future additional capabilities could make the MDT technique an even better evaluation tool.

Regardless of the MDT method's ability to perform the tasks of evaluations, its use in American education will depend a great deal upon its acceptability to students and instructors. For this reason, the above research is highly important to provide both instructors and students with an understanding that this testing method is acceptable when used appropriately. Specifically, the above reported research on students' attitudes, when controlled for the factor of appropriate usage, should be especially useful to encourage other instructors to utilize the method with confidence. The MDT method is demonstrated to be perceived by students as an acceptable step forward in the offering of different and more rigorous alternatives for educational testing.

BIBLIOGRAPHY

Anderson, Paul S. (1987a) The MDT Innovation: Machine Scoring of Fill-in-the-Blank. Multi-Digit Technologies Corporation, Normal, IL pp. 190.

Anderson, Paul S. (1987b) "Student Attitudes about MDT Multi-Digit Testing: Analyses from Pioneer Experiences," Paper at the annual conference of the National Council for Meassurement in Education, Washington, D.C., April 19-23, 1987.

APPENDIX A:

SURVEY OF STUDENT OPINIONS ABOUT METHODS OF EDUCATIONAL TESTING

Please answer these questions on the new MDT answer sheet (F3). Note that it has Short Answer SA (Essay) spaces at the bottom to make written comments to elaborate on the encoded responses.

START on QUESTION 21 on back of the answer sheet.

A. In your **high school education**, **how much experience** did you have with each of these test methods?

Question No.		Almost None;	Little;	Some;	Much;	Very Much;
21.	True/False	1	2	3	4	5
22.	Multiple Choice	1	2	3	4	5
23.	Matching	1	2	3	4	5
24.	MDT Multi-Digit	1	2	3	4	5
25.	Fill-in-the-blank	1	2	3	4	5
26.	Short answer (sentence +)	1	2	3	4	5
27.	Essay (paragraph +)	1	2	3	4	5

B. In your **university education**, how much experience have you had with each of these test methods?

Question No.		Almost None;	Little;	Some;	Much;	Very Much;
28.	True/False	1	2	3	4	5
29.	Multiple Choice	1	2	3	4	5
30.	Matching	1	2	3	4	5
31.	MDT Multi-Digit	1	2	3	4	5
32.	Fill-in-the-blank	1	2	3	4	5
33.	Short Answer (sentence +)	1	2	3	4	5
34.	Essay (paragraph +)	1	2	3	4	5

C. **Rate your ability as a test taker** in each of the following methods of testing. (Note: This is NOT a ranking; you could be poor or good at all.)

Question No.		Very Poor;	Poor;	Average;	Good;	Very Good;
35.	True/False	1	2	3	4	5
36.	Multiple Choice	1	2	3	4	5
37.	Matching	1	2	3	4	5
38.	MDT Multi-Digit	1	2	3	4	5
39.	Fill-in-the-Blank	1	2	3	4	5
40.	Short Answer (sentence +)	1	2	3	4	5
41.	Essay (paragraph +)	1	2	3	4	5

D. Based upon your test experiences, please **rate these test methods according to how well they can evaluate student learning.**:

Question No.		Very Poorly;	Poorly;	Average;	Well;	Very Well;
42.	True/False	1	2	3	4	5
43.	Multiple Choice	1	2	3	4	5
44.	Matching	1	2	3	4	5
45.	MDT Multi-Digit	1	2	3	4	5
46.	Fill-in-the-Blank	1	2	3	4	5
47.	Short Answer (sentence +)	1	2	3	4	5
48.	Essay (paragraph +)	1	2	3	4	5

E. **In general**, what is your attitude about each method of testing?

Question No.		Strongly Dislike;	Dislike;	Neutral;	Like;	Strongly Like
49.	True/False	1	2	3	4	5
50.	Multiple Choice	1	2	3	4	5
51.	Matching	1	2	3	4	5
52.	MDT Multi-Digit	1	2	3	4	5
53.	Fill-in-the-Blank	1	2	3	4	5
54.	Short Answer (sentence +)	1	2	3	4	5
55.	Essay (paragraph +)	1	2	3	4	5

56. Would you recommend the continued use of the MDT testing method in this course? 1. strongly "no"; 2. basically "no"; 3. neutral; 4. basically "yes"; 5. strongly "yes"

57. Would you recommend the use of the MDT method for any other courses? 1. strongly "no"; 2. basically "no"; 3. neutral; 4. basically "yes"; 5. strongly "yes"

58. Do you consider the MDT method to be a valid or invalid way of testing when applied to the learning of discrete facts? 1. highly invalid; 2. moderately invalid; 3. neutral; 4. moderately valid; 5. highly valid

59. If given the option to enroll in either of two sections of another course, knowing that one would use the MDT method and the other would not, what would be your choice? 1. Definately avoid the MDT method, even if you had to adversely adjust your schedule of other classes; 2. Try to avoid the MDT method if class schedule permits; 3. Neutral, it makes no difference; 4. Try to enroll in the MDT section if class schedule permits; 5. Definitely enroll in the MDT section even if you had to adversely adjust your schedule of other classes.

60. In comparison with studying for mutiple choice and fill-in-the-blank questions, how should a student prepare for MDT Multi-Digit questions on a test? 1. The same as for multiple choice questions; 2. The same as for fill-in-the-blank questions; 3. Just study normally because the three test methods are all so similar; 4. Altogether differently (please comment in the SA space on the answer sheet).

NOTE: For research purposes of comparisons and follow-up, mark your name and Social Security Number on the answer sheet. Your data will be confidential.

Please continue with the questions 1-20. These questions are answered on the front (Multi-Digit) side of the answer sheet. You are almost finished.

Question No.

1. What is your **sex**? 001=male; 002=female.

2. What is your **class status**? 001=freshman; 002=sophomore; 003=junior; 004= senior; 005=graduate; 006=other.

3. What is your **age**? (Encode the actual years. For example, if you are 21, encode 021.)

4. What is your **major** (or probable major)? 001=teacher education/special education; 002=social sciences; 003=fine arts/languages; 004=physical sciences/math; 005=computer/applied technology; 006=business management, accounting, marketing, etc.; 007=truly undecided. Please also write your major (or probable major) in space SA101 at the bottom of the answer sheet.

5. How closely does this course relate to your major and intended future employment? 001=Not at all; 002=very little; 003=some; 004=reasonable amount; 005 very much.

6. What is your **overall GPA** at ISU? 001=less than 1.75; 002=1.75 to 1.99; 003=2.00 to 2.24; 004=2.25 to 2.49; 005=2.50 to 2.74; 006=2.75 to 2.99; 007=3.00 to 3.24; 008=3.25 to 3.49; 009=3.50 to 3.74; 010=3.75 to 4.00.

7. Overall, how would you rate your instructor in this course? 001=bad; 002=poor; 003=average or okay; 004=good; 005=excellent.

8. Please classify yourself as an ISU student in terms of effort. 001=very low; 002=lower than most; 003=medium; 004=higher than most; 005=very high.

9. Please classify yourself as an ISU student in terms of natural intelligence (ability). 001=very low; 002=lower than most; 003=medium; 004=higher than most; 005=very high.

10. What grade do you expect to receive in this course? 001=F; 002=D/F; 003=D; 004=D/C; 005=C; 006=C/B; 007=B; 008=B/A; 008=B/A; 009=A.

11. What grade do you think you deserve in this course (based on effort and what you have learned during this semester)? 001=F; 002=D/F; 003=D; 004=D/C; 005=C; 006=C/B; 007=B; 008=B/A; 009=A.

12. How much "prior knowledge" of the subject matter did you have before taking this course? 001=none; 002=very little; 003=little; 004=some; 005=much; 006=very much; 007=almost all.

13. Counting this course, **how many courses** at ISU have you had with tests using the MDT method? **Code in the actual number.** (For example, three courses would be 003.) Also, **please name** them in the space SA102 for written comments on the answer sheet.

14. Counting this course, how many of those courses using the MDT method are during this Fall 1986 semester? **Code in the actual number.** Also, please circle them in SA102.

15. In total for all your courses ever at ISU, how many tests have you taken with MDT style questions?

16. Are the MDT testing procedures as used in this course appropriate for the course material? Mark your answer and then please comment in the SA space on the answer sheet. 001=very inappropriate; 002=inappropriate; 003=appropriate; 004=highly appropriate.

17. Are the other testing procedures as used in this course appropriate to the course material? (Please comment and/or suggest alternatives.) 001=very inappropriate; 002=inappropriate; 003=appropriate; 004=highly appropriate.

18. Are you being graded fairly in this class? 001=very fairly; 002=unfairly; 003=average/fairly; 004=very fairly.

Please comment in the SA spaces on the answer sheet. We read your comments.

Please be sure that you have answered all of the questions. Incomplete data is unnecessarily difficult to analyze. Thank you for your cooperation.